

ABSTRACT OF THE DISCLOSURE

A system and method for detecting file content similarity within a file system. The system may include a storage device configured to store a plurality of files and a file
5 system configured to manage access to the storage device. The file system may be configured to compute a compressed size of at least a portion of a given file, and to store an indication of the compressed size in a named stream corresponding to the given file. In one specific implementation of the system, the file system may be further configured to determine a value of a file harmony metric from the compressed size of the concatenation
10 of a first and second file and the compressed sizes of the first and second files, where the compressed sizes of the first and second files may be stored in a corresponding one of a plurality of respective named streams.